1 Fraud

Charlie's Angels

2024-05-18

```
knitr::opts_chunk$set(warning = FALSE, message = FALSE)
```

Libraries to be used:

```
library(readr)
library(tidyr)
library(dplyr)
library(lubridate)
library(ggplot2)
library(magrittr)
library(data.table)
```

A. Importing Fraud Data

```
read_fraud <- function(folder = getwd()) {
  files <- list.files(path = folder, pattern = "*.csv", full.names = TRUE)
  transactions_list <- lapply(files, fread)
  transactions <- bind_rows(transactions_list)
  transactions <- transactions %>%
    mutate(
      CardType = as.factor(CardType),
      Fraud = as.factor(Fraud)) %>%
    select(TimeStamp, CardType, Amount, Fraud)
return(transactions)
}

#Input your directory here (to verify)
transactions <- read_fraud("C:/Users/amore_6ou078y/OneDrive/Documents/UP Subjects/Stat 125/R/Fraud Data
head(transactions)
```

```
##               TimeStamp CardType  Amount  Fraud
##                  <POSc>   <fctr>   <num> <fctr>
## 1: 2023-01-01 00:00:05       Dr  640.28     No
## 2: 2023-01-01 00:00:18       Cr 1500.00    Yes
## 3: 2023-01-01 00:00:25       Cr 3821.77     No
## 4: 2023-01-01 00:00:44       Cr 4849.85     No
## 5: 2023-01-01 00:00:47       Dr 1500.00    Yes
## 6: 2023-01-01 00:00:57       Dr  220.19     No
```

B. Single Line Codes

```
loss_tally <- filter(.data = transactions, Fraud == "Yes") %>% summarise(Sum = sum(Amount))
loss_tally
```

```
##        Sum
## 1 39465197
```

1. For the whole year of 2023, how much did the bank lose due to fraud transactions?

- *39,465,197* pesos lost due to fraud transactions.

```
date_transactions <- filter(.data = transactions, Fraud == "Yes") %>%
  mutate(Date = as.Date(TimeStamp)) %>%
  group_by(Date, Fraud) %>%
  summarise("number" = n(), .groups = 'drop') %>%
  arrange(desc(number))

head(date_transactions, 4)
```

```
## # A tibble: 4 x 3
##   Date       Fraud number
##   <date>     <fct>  <int>
## 1 2023-12-24 Yes      336
## 2 2023-12-31 Yes      333
## 3 2023-12-25 Yes      323
## 4 2023-01-01 Yes      219
```

2. Find the top 4 days with the greatest number of fraudulent transactions.

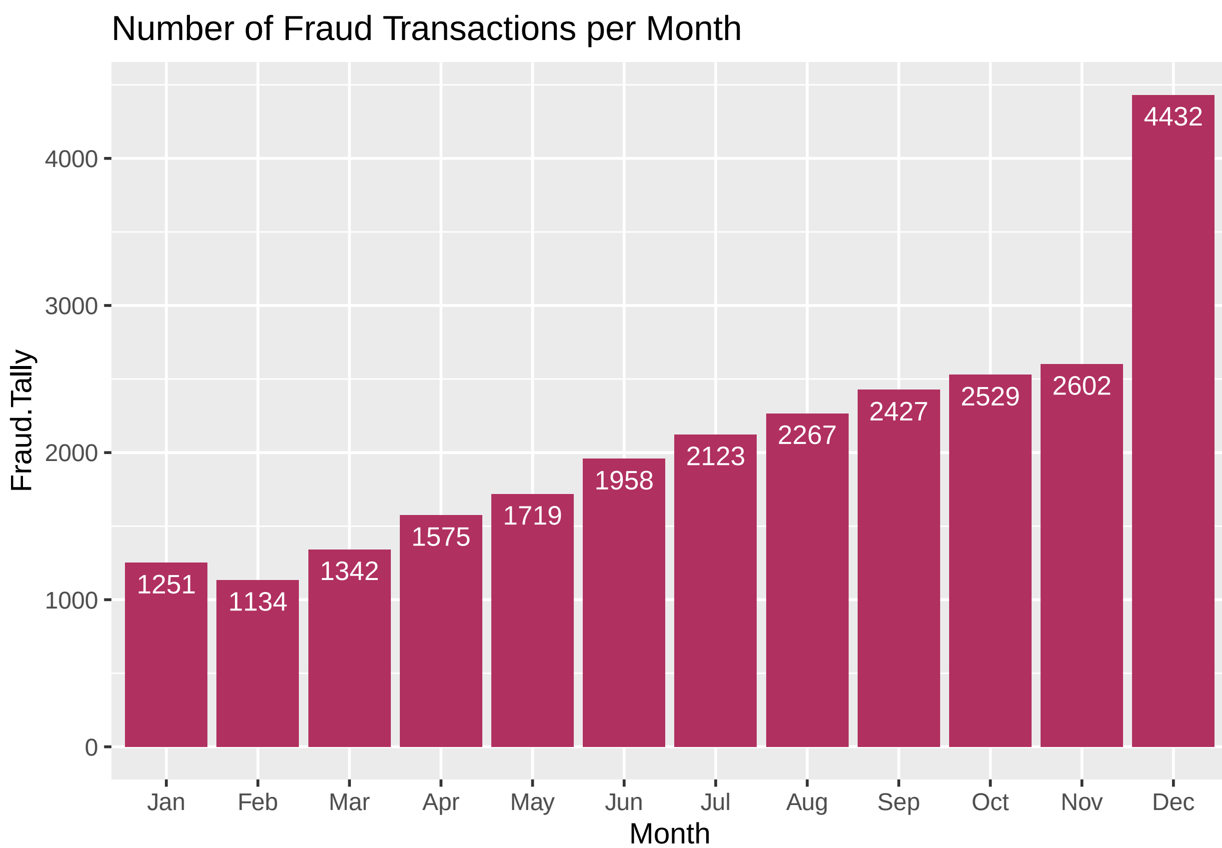
1. December 24, 2023 (*336*)
2. December 31, 2023 (*333*)
3. December 25, 2023 (*323*)
4. January 1, 2023 (*219*)

---

3. Create a bar chart showing number of fraud transactions per month.

```
by_month <- mutate(date_transactions, Month = lubridate::month(Date, label = TRUE)) %>%
  group_by(Month) %>%
  summarise("Fraud.Tally" = sum(number)) %>%
  ggplot(aes(x = Month, y = Fraud.Tally)) +
  geom_bar(stat="identity", fill = 'maroon')+
  geom_text(aes(label=Fraud.Tally), vjust=1.6, color="white", size=3.5) +
  ggtitle("Number of Fraud Transactions per Month")

by_month
```



```
by_CardType <- select(transactions, CardType, Fraud) %>%
  mutate(value = ifelse(Fraud == "Yes", 1, 0)) %>%
  group_by(CardType) %>%
  summarise("number" = n(), tally = sum(value)) %>%
  mutate("P(fraud|CardType)" = tally/number)

by_CardType
```

```
## # A tibble: 2 x 4
##   CardType  number tally `P(fraud|CardType)`
##   <fct>      <int> <dbl>               <dbl>
## 1 Cr       1622845  1966             0.00121
## 2 Dr       3785633 23393             0.00618
```

4. Which type of card is less prone to fraud: credit or debit cards?

- **P(fraud | Credit)** = *0.001211453*
- **P(fraud | Debit)** = *0.006179416*

  Thus, **Credit cards are less prone to fraud**, since relative frequency of fraud in Credit cards is lower than in Debit cards.
